DOCUMENT RESUME

ED 407 418 TM 026 440

AUTHOR Gray, B. Thomas
TITLE Higher-Order Factor Analysis.
PUB DATE 24 Jan 97
NOTE 22p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Austin, TX, January 23-25, 1997).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Correlation; *Factor Analysis; *Matrices; *Orthogonal Rotation
IDENTIFIERS *Second Order Effects

ABSTRACT

Higher order factor analysis is an extension of factor analysis that is little used, but which offers the potential to model the hierarchical order often seen in natural (including psychological) phenomena more accurately. The process of higher order factor analysis is reviewed briefly, and various interpretive aids, including the Schmid-Leiman solution, are discussed. An example of the use of higher-order factor analysis is provided using the Alcohol Use Inventory. The basic process of factor analysis can be conceptualized in terms of a series of matrices. A matrix of data is analyzed to produce a matrix of associations. An appropriate extraction technique is used to produce the factor matrix. An interfactor matrix of associations (factors by factors) is constructed, and factors are again extracted to yield higher order factors that can be rotated. Repeating the process will yield sequentially higher-order factors until either a single factor is extracted, or the extracted factors are uncorrelated even with rotation. Interpreting the higher order factor follows. The solution proposed by J. Schmid and J. Leiman (1975)"orthogonalizes" the factors by residualizing the variance from the primary factors and attributing it to the second-order factor alone. This approach gives another look at a data set that may provide useful information. Higher-order factor analysis is not often used, but it has the potential to aid interpretation. (Contains 1 figure, 6 tables, and 29 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 407 418

Higher-Order Factor Analysis

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
B. THOMAS GRAY
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

B. Thomas Gray

Texas A&M University 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX, January 24, 1997.

TM026440

BEST COPY AVAILABLE

Abstract

Higher-order factor analysis is an extension of factor analysis that is little used, but which offers the potential to more accurately model the hierarchical order often seen in natural (including psychological) phenomena. The process of higher order factor analysis is briefly reviewed, and various interpretive aids, including the Schmid-Leiman solution, are discussed. An example of the use of higher-order factor analysis is provided using the Alcohol Use Inventory.

Higher-Order Factor Analysis

Factor analysis is a technique that allows for the reduction of a data set with a large number of variables to one with a smaller, and therefore more manageable number of factors. As Gall, Borg, and Gall (1996) noted, "Factor analysis provides an empirical basis for reducing all these variables to a few factors by combining variables that are moderately or highly correlated with each other" (pp. 447-448). The researcher is thus provided with a set of information which must then be interpreted in a theoretically consistent fashion. As Gorsuch (1983, p. 2) has stated, "Usually the aim [in using factor analysis] is to summarize the interrelationships among the variables in a concise but accurate manner as an aid to conceptualization."

Many phenomena, including psychological occurrences, are conceptualized as being hierarchically ordered. For example, Gorsuch (1983) discussed the way the Earth's topography is frequently categorized, being first divided according to land or water. Landforms are subdivided into continents and islands, and each can be further subdivided by location. Similarly, bodies of water are divided into oceans, lakes and streams.

If we conceptualize nature as consisting of hierarchically-ordered phenomena, then it is only logical to model the phenomena in this way. This is reflected in the structure of psychological tests, which often include several levels of subtests or subscales, and are therefore implicitly hierarchical. One example is the Wechsler tests of intelligence, which subdivide g, or general intelligence, into verbal and performance domains. The examinee's score for either of these domains may be further broken down according to the subtests of which they are composed.

The essential concept of higher order factor analysis follows this same line of reasoning. Factors have been conceptualized as groupings of variables that share an acceptable amount of

variance, or in other words, variables that are correlated with one another. Higher order factors similarly are groupings of factors that are more closely correlated with on another than they are to other factors or factor groupings. The process involves iterations of extracting higher-order factor(s) from the relevant lower-order interfactor matrix of associations until either only a single factor is derived (for example, g) or until the lower- and higher-order factors are the same. This process is explained in greater detail below.

A recent review of factor analytic studies in the field of counseling psychology (Tinsley & Tinsley, 1987) failed to mention of higher-order factor analysis, which would seem to bear out Kerlinger's (1984, p. xivv) comment that the procedure "seems not to be widely known or understood." Indeed, it is not typically mentioned in the sections of texts which include substantial introductions to factor analysis (e.g., Crocker & Algina, 1986; Stevens, 1996). This is disturbing, and given the promise that higher-order factor analysis has in terms of reflecting the hierarchical nature of many naturally occurring phenomena, it is to be hoped that future workers will become more familiar with the technique, and consider its use more frequently.

Brief Review of the Factor Analytic Process

The basic process of factor analysis can be readily conceptualized in terms of a series of matrices, as portrayed in Figure 1 (see Hetzel, 1995, for a very readable description of basic factor analysis; a more detailed treatment is to be found in Gorsuch, 1983). A matrix of data (**X**, items by variables) is analyzed to produce a matrix of associations (**R**, variables by variables), usually by either computing correlation coefficients between the different variables or developing the relevant variance-covariance matrix. An appropriate extraction technique, such as principal components analysis (PCA, principle factors analysis, or PFA, is an alternative technique) is then used to produce the factor

matrix (**F**, variables by factors). Any method of extraction will produce factors which are orthogonal, and the matrix of associations between the factors will therefore by definition be an "identity" matrix. Efforts to extract factors from such a matrix will not yield new factors, but will instead simply reproduce the original set of factors. It is then possible to rotate the matrix obliquely, which will redistribute the variance such that the factors are now correlated (the resultant matrix is labeled **F'**).

The process just described produces primary factors, and the use of oblique rotation implies that they are correlated. Given the nature of the variables most often employed in behavioral science research, correlated variables are to be expected, since investigation of a given construct will almost invariably involve measurements which tap into the same part(s) of the construct. It is therefore reasonable to expect that the variables might be hierarchically related to one another, which in turn would make use of higher-order factor analysis appropriate.

Continuing the process outlined above results in the extraction of second order factors (Fig. 1). An interfactor matrix of associations (**R**, factors by factors) is constructed, and factors are then extracted from it using PCA (or PFA), or another suitable method (Gorsuch, 1983). The resultant higher-order factor matrix (**H**, factors by higher-order factors) can then be rotated. Repeating the process will yield sequentially higher-order factors until either only a single factor is extracted, or until the factors extracted are uncorrelated even with rotation.

One notable difference from first-order factor analysis is that the statistical significance of the matrix (Bartlett, 1950) cannot be used as a test to determine the number of factors to retain in a higher-order analysis. This is because the sampling distribution of correlation coefficients will in part be a function of whatever rotation strategy is employed, and thus the distribution will vary according to rotation strategy (Gorsuch, 1983). In fact, this is not a great loss anyway, since the utility of

statistical significance testing is limited, at best (see Cohen, 1995; Thompson, 1989, 1994, 1996).

Interpretation of Higher-order Factors

The next task facing the researcher is to make sense of the higher-order factor, which is to say, to interpret its meaning. An approach is to base the interpretation of the higher-order factor on the interpretations of the lower-order factor(s) from which it is derived, as Kerlinger (1984) did in his study of social attitudes (see Thompson, 1985). This may be superficially appealing, but it must be borne in mind that any such interpretation necessarily involves elimination of some information. Certainly the purpose of analyses is to remove the information (i.e., the variance) that is not useful for explaining the phenomenon of interest. However, as Gorsuch (1983, p. 245) pointed out, this amounts to "basing interpretations based on interpretations." Information that is deleted at one step is potentially relevant at the next.

Several solutions have been offered to help in resolving this problem. All provide the researcher with a variable-by-higher-order-factor matrix of factor pattern coefficients, although they are derived by different methods. Gorsuch (1983) suggested that the primary factor pattern matrix be postmultiplied by the higher-order factor pattern matrix ($\mathbf{P}_{vf}\,\mathbf{P}_{fh} = \mathbf{P}_{vh}$). Thompson (1990) carried this a step further by applying a Varimax rotation to the resultant product matrix ($\mathbf{P}_{vh}$), which seems more consistent with the procedures employed in these analyses.

Schmid and Leiman (1957) offered a slightly different approach. Their procedure follows the usual process, but the final result distributes the variance somewhat differently. The Schmid-Leiman solution "orthogonalizes" the factors by residualizing the variance from the primary factors and attributing it to the second-order factor alone. Borrello and Thompson (1990) applied this method in testing the validity of Lee's (1973/1976) typology of love as formulated by Hendrick and Hendrick

(1986). The six basic types of love defined by Lee appeared as primary factors, but the predicted pattern of relationships among them was not supported by the second-order factors which emerged. Noteworthy was the fact that the six basic types were discernible at all steps in the analysis, even with the variance common to the first- and second-order factors removed from the former (Borrello & Thompson, 1990).

Thompson (1990) has suggested that elucidating first- and second-order factors from a data set is analogous to looking at a mountain range from a close-up view and again from further away. Following the same line of thinking, it is here suggested that the initial matrix of associations derived from a large data set might be somewhat like being on the streets of New York City, while the primary factors would be like viewing it from atop the Empire State Building. A second-order factor analysis would give the perspective of an airline pilot flying over the city, and a third-order analysis would be like the view from the space shuttle.

SECONDOR

Thompson (1990) has developed a FORTRAN program that greatly facilitates analysis using the strategy of higher-order factors. The output generated by SECONDOR includes several sets of results. A first order PCA is provided, including an unrotated solution, as well as the Varimax and Promax rotated solutions. The program also allows the worker the option of either retaining all factors with eigenvalues greater than 1, or else manually selecting the number of factors to retain. In addition, SECONDOR provides both an unrotated and a Varimax rotated second-order solution, and a Schmid-Leiman solution. It was this program that was used to generate the output used in the analysis of the love typology as briefly discussed above (Borrello & Thompson, 1990).

An Illustrative Example: The Alcohol Use Inventory (AUI)

The AUI is a 147 item instrument designed to evaluate an examinee's patterns of alcohol use (Horn, Wanberg, & Foster, 1974; Wanberg, Horn, & Foster, 1977). It has been thoroughly studied by Skinner (1981a, 1981b; Skinner & Allen, 1983). As reported in Table 1, sixteen scales have been defined, based on prior factor analytic studies, two separate and independent samples yielded mean internal consistency reliability indices on the order of .75 (Skinner, 1981a; Wanberg, et al., 1977). Wanberg and his colleagues (1977) conducted an analysis of the data from 2,261 administrations of the instrument over a period of about four years. The correlation matrix of the 16 scales for this sample is presented in Table 2.

Unfortunately, Wanberg et al. (1977) failed to provide adequate details of the procedure they followed to be able to duplicate their analysis (see Skinner, 1983, for discussion of the general failure to adequately report factor analytic studies pertaining to alcohol misuse). The correlation matrix of Table 2 was analyzed using the SECONDOR program. Eigenvalues for the first six factors derived through the principal components analysis were 4.999, 1.518, 1.445, 1.334, 0.951, and 0.892. Given the fairly large separation between the fourth and fifth values, the Guttman rule was applied, and the first four factors were retained. The factor pattern matrix for the Varimax-rotated solution is presented in Table 3. The Promax factor pattern and structure pattern matrices yield essentially the same results.

While there are some overall similarities between the analysis reported here using SECONDOR, and the previous work of Wanberg and his associates (1977), there are also some notable differences. Factor I in Table 3 essentially corresponds to Factor D_1 (Deterioration) of Wanberg, et al., and Factor IV to their Factors A and B together (save that the factor pattern and factor structure coefficients for Variable 3 are relatively small in Factor IV). Factor II in the present

analysis includes the same variables as Factor C of Wanberg, et al., and also includes Variables 15 and 16. There is no apparent correlate in the scheme of Wanberg, et al. to Factor III. It is not surprising that there is nothing in the SECONDOR analysis corresponding to Factor D_2 of Wanberg, et al., since the latter is composed of some of the 25 questions not included in the 16 scales that form the variables for this study.

A General Alcoholism (g) Factor was also identified by Wanberg and associates (1977), although again the basis for this, including its variable composition, were unreported. SECONDOR yielded two second-order factors in the present analysis, and the Varimax-rotated solution is presented in Table 4. This would seem to indicate that the second-order factor labeled H1 in Table 4 is made up of primary factors I, II, and IV; and that H2 is then made up only of primary factor III. However, Table 5 presents the Varimax-rotated product matrix of the primary and the second-order factor pattern matrices, and examination of this data shows clearly that the variables do not sort into higher-order factors as neatly as Table IV might imply. Indeed, these would seem to represent rather distinct constructs, whose interpretation is better left to another arena.

Table 6 presents the Schmid-Leiman solution for this same data. This illustrates that the "orthogonalized" first-order factors retain the same basic composition as that seen in Table 3, and again that the higher-order factors are not composed simply of combinations of the primary factors.

Summary

The above discussion has presented the process of higher-order factor analysis. Various interpretation aids have been reviewed, and illustrative examples have been provided. Given this information, it is troubling that relatively few applications have been made of this analytic tool (cf. Tinsley & Tinsley, 1986). Of particular concern is the commentary of Nunnally (1978, pp. 431-432),

who argued against the use of higher-order factor analysis on two grounds:

> The average psychologist has difficulty in understanding first-order factors, and this difficulty is increased with higher-order factors... Also, if factor analysis is partly founded on the principle of parsimony, it is reasonable to question the parsimony of having different orders of factors.

To the first claim, that it is just "too difficult," it might be suggested that Nunnally underestimated the abilities of many psychologists.

The question of parsimony is equally puzzling. Certainly the use of higher-order factor analysis requires a greater expenditure of time and effort. What is gained by this, however, is a greater wealth and diversity of information from a given data set. Higher-order analyses offer the ability to simplify information in ways which potentially aid interpretation, and consequently the ability to more fully understand the phenomenon of interest. It is hoped that future workers will be more willing to put forth the energy necessary for these analyses.

References

Bartlett, M. S. (1950). Tests of significance in factor analysis. British Journal of Psychology, 3(2), 77-85.

Borrello, G. M., & Thompson, B. (1990). An hierarchical analysis of the Hendrick-Hendrick measure of Lee's typology of love. Journal of Social Behavior and Personality, 5, 327-342.

Cattell, R. B. (1965). Higher order factor structures and reticular-vs-hierarchical formulae for their interpretation. In C. Banks & P. L. Broadhurst (Eds.) ΣΤΕΦΑΝΟΣ: Studies in psychology presented to Cyril Burt (pp. 223-266). New York: Barnes & Noble (American Edition published 1966 Univ. of London Press).

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. New York: Harcourt Brace Jovanovich.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). Educational research: An introduction (6th ed.) White Plains, NY: Longman.

Gorsuch, R. L. (1983). Factor analysis (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Hendrick, C., & Hendrick, S. S. (1986). A theory and method of love. Journal of Personality and Social Psychology, 50, 392-402.

Hetzel, R. D. (1995, April). The use of factor analysis in counseling psychology research: Current practice is not necessarily best practice. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Horn, J. L., Wanberg, K. W., & Foster, F. M. (1974). The Alcohol Use Inventory. Denver, CO: Center for Alcohol Abuse Research and Evaluation.

Kerlinger, F. N. (1984). Liberalism and conservatism: The nature and structure of social attitudes. Hillsdale, NJ: Lawrence Erlbaum.

Lee, J. A. (1973). The colors of love: An exploration of the ways of loving. Don Mills, Ontario: New Press. (Popular Edition, 1976).

Nunnally, J. C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.

Schmid, J., & Leiman, J. (1957). The development of hierarchical factor solutions. Psychometrika, 22, 53-61.

Skinner, H. A. (1980). Factor analysis and studies on alcohol: A methodological review. Journal of Studies on Alcohol, 41, 1091-1101.

Skinner, H. A. (1981a). Primary syndromes of alcohol abuse: Their measurement and correlates. British Journal of Addiction, 76, 63-76.

Skinner, H. A. (1981b). Comparison of clients assigned to in-patient and out-patient treatment for alcoholism and drug addiction. British Journal of Psychiatry, 138, 312-320.

Skinner, H. A., & Allen, B. A. (1983). Differential assessment of alcoholism: Evaluation of the Alcohol Use Inventory. Journal of Studies on Alcohol, 44, 852-862.

Stevens, J. (1996). Applied multivariate statistics for the social sciences (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Thompson, B. (1985). Review of Liberalism and conservatism: The nature and structure of social attitudes by F. N. Kerlinger. Educational and Psychological Measurement, 45, 429-430.

Thompson, B. (1989). Statistical significance, result importance, and result generalizability: Three noteworthy but somewhat different issues [Editorial]. Measurement and Evaluation in Counseling and Development, 22, 2-5.

Thompson, B. (1990). SECONDOR: A program that computes a second-order principal components analysis and various interpretation aids. Educational and Psychological Measurement, 50, 575-580.

Thompson, B. (1994). The concept of statistical significance testing. Measurement Update, 4(1), 5-6.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. Educational Researcher, 25(2), 26-30.

Wanberg, K. W., Horn, J. L., & Foster, F. M. (1977). A differential assessment model of alcoholism: The scales of the Alcohol Use Inventory. Journal of Studies on Alcoholism, 38, 512-538.

Table 1

Scales of the Alcohol Use Inventory (AUI)

1	Drink to Improve Sociability - Social Benefit
2	Drink to Improve Mental Functioning - Mental Benefit
3	Gregarious versus Solitary Drinking
4	Obsessive-Compulsive Drinking
5	Continuous, Sustained Drinking
6	Postdrinking Worry, Fear and Guilt
7	Drink to Change Mood
8	External Support to Stop Drinking
9	Loss of Behavior Control when Drinking
10	Social-Role Maladaptation
11	Psychoperceptual Withdrawal
12	Psychophysical Withdrawal
13	Nonalcoholic Drug Use
14	Quantity of Alcohol Used
15	Drinking Followed Marital Problems
16	Drinking Provokes Marital Problems

Table 2

Correlations between 16 AUI Scales

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	59	1.00														
3	31	13	1.00													
4	31	33	-.02	1.00												
5	15	20	06	39	1.00											
6	37	19	03	43	09	1.00										
7	49	33	10	35	10	56	1.00									
8	11	09	-.13	29	-.04	23	22	1.00								
9	27	16	10	42	04	49	38	24	1.00							
10	26	18	19	35	01	25	23	20	44	1.00						
11	26	24	09	49	13	32	31	32	54	51	1.00					
12	33	25	11	55	15	49	39	31	55	46	62	1.00				
13	13	10	18	16	10	07	12	12	24	22	23	17	1.00			
14	25	17	28	38	10	20	19	14	44	53	50	46	27	1.00		
15	20	13	14	11	00	13	31	03	21	21	20	16	13	20	1.00	
16	24	16	08	20	-.03	44	33	08	43	21	17	25	07	16	25	1.00

Note. From "A differential assessment model of alcoholism: The scales of the Alcohol Use Inventory," by K. W. Wanberg, J. L. Horn, and F. M. Foster, 1977, Journal of Studies on Alcoholism, 38, p. 521. Copyright 1977 by the Journal of Studies on Alcohol Inc., Rutgers Center of Alcohol Studies, Piscataway, NJ 08855. Reprinted with permission.

Table 3

First-Order Varimax Matrix and h^2

	First-Order Factors				
Variable	I	II	III	IV	h^2
1	.141	.501	.279	.593	.700
2	.066	.292	.167	.692	.597
3	.302	.085	.708	.137	.619
4	.460	.191	-.395	.543	.698
5	.089	-.217	-.090	.719	.580
6	.176	.700	-.306	.204	.655
7	.097	.721	-.069	.332	.644
8	.290	.177	-.531	.024	.398
9	.592	.503	-.186	.000	.638
10	.733	.203	.039	.010	.581
11	.733	.198	-.233	.183	.665
12	.624	.345	-.291	.256	.658
13	.492	-.031	.206	.074	.291
14	.798	.061	.132	.091	.667
15	.212	.454	.317	-.066	.356
16	.106	.727	.009	-.091	.548

Table 4

Varimax-Rotated Second-Order Solution and h^2

First-Order Factor	Second-Order Factors		
	H1	H2	h^2
I	.805	-.076	.653
II	.772	.179	.628
III	-.042	.970	.942
IV	617	-.208	.424

Table 5

Varimax-Rotated Product Matrix and h^2

	Second-Order Factors		
Variable	H1	H2	h^2
1	.664	-.277	.518
2	.457	-.316	.308
3	.662	.285	.520
4	.182	-.793	.662
5	.005	-.388	.151
6	.346	-.577	.453
7	.501	-.440	.444
8	-.098	-.565	.329
9	.444	-.526	.474
10	.472	-.353	.348
11	.334	-.628	.506
12	.352	-.695	.607
13	.345	-.096	.128
14	.492	-.316	.343
15	.552	.045	.307
16	.468	-.189	.255

Table 6

Schmid-Leiman Solution for AUI Data

	Second-Order Factors		First-Order Factors				
Variable	H1	H2	I	II	III	IV	h^2
1	.667	.271	-.023	.286	.075	.431	.792
2	.546	.097	-.056	.160	.046	.527	.616
3	.269	.669	.171	-.012	.179	.102	.592
4	.687	-.435	.231	.053	-.080	.359	.853
5	.277	-.272	.020	-.172	-.015	.573	.509
6	.652	-.166	.015	.437	-.067	.082	.655
7	.665	.040	-.052	.454	-.011	.198	.692
8	.328	-.470	.169	.086	-.120	-.040	.381
9	.686	-.061	.328	.247	-.029	-.098	.653
10	.584	.081	.454	.022	.028	-.072	.560
11	.680	-.211	.440	.022	-.036	.055	.705
12	.740	-.246	.344	.135	-.052	.110	.759
13	.312	.175	.317	-.102	.062	.024	.243
14	.572	.122	.505	-.085	.052	.000	.608
15	.361	.421	.088	.260	.082	-.092	.398
16	.465	.196	-.009	.465	.005	-.134	.489

Figure 1

Diagramatic Representation of the Factor Analytic Process
(including Extraction of Higher-Order Factors)